Print This Page

Close This Window

Team Control Number

41224

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

Problem Chosen

A

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2015 Mathematical Contest in Modeling (MCM) Summary Sheet
(Attach a copy of this page to your solution paper.)

Type a summary of your results on this page. Do not include
the name of your school, advisor, or team members on this page.

text

[TITLE]

1. Introduction

Academic success among undergraduate students at universities in the U.S. depends on several factors such as teacher capability, class size, and university funding. In nearly every statistical study regarding factors that contribute to student success, the amount of funding per student at each school always held the greatest positive impact.[**?**] While great teachers and small class sizes certainly effect student success in the classroom, we know that adequate funding is essential to resource expansion in U.S. schools.[**?**] Additional resources that are readily available to the student have been shown to increase student performance across the board.

If a foundation should want to increase student success by providing funds to several universities, the foundation could pursue one of several options.
Most grant-awarding foundations that are prominent today allocate funds according to a recipients's qualifications. These qualifications are usually determined via a selection process that involves an application or proposal from the university, followed by a direct review of that university by the foundation. Successful candidates for grants usually demonstrate adequate need for the award, as well as a plan to use the money to ameliorate the situation at the school. Universities with both need and potential for growth are selected for grants primarily by human opinion and group decision. (unsure how to cite this paragraph, information came from browsing various foundation websites)

In order to avoid this somewhat meticulous selection process, we develop a model that will rank and allot appropriate funds to each university for the foundation. The model employs methods of data clustering, neural networks, and multi-objective programming to essentially replace the human-performed decision making process of previously used grant allocation techniques.
We propose a model that will rank universities according to their eligibility for the grant, amount of grant money to be received, and rate at which the grant money will be distributed.

+ The model ranks all universities according to their current available funds due to outside donations each year in addition to how each university employs those funds. Universities with relatively small donation pools and a loyal history of fund allocation to expansion of student resources receive high rankings.
+ The model approximates the amount of effective change it can induce by giving funds to schools of various rankings. The schools with optimized rates of changes will receive the largest grants from the Foundation.
+ The size of grant awarded to each school in the list will be determined by XXXXXXX.
+ Over a period of five years, the foundation will award each grant according to a predictive distribution of optimal funding per year at each institution. This data comes from published financial data regarding university donation spending.

2. Assumptions

These are some things we assumed in order to create our model!

3. Formal Prolegomena

First, some notational definitions. We have tabulated them below.

($\mathcal{U}$) The set of universities and colleges in question.

($\mathcal{D}$) Financial space. In general, this is the space of donations – this might have multiple dimensions over $\mathbb{R}$, depending on the specific categories of money we're interested in. It might also contain certain categories of expenditures that we might be able to influence by placing constraints on how institutions are allowed to spend their money.

($\mathcal{T}$) The space of times for which we have data.

($\mathcal{V}$) The vector space of student metric variables $\{v_i\}$. Note that at this point, we have not yet committed ourselves to any such choice of variables, and so $\mathcal{V}$ includes also negative and neutral indicators of success.

($\frac{d\mathcal{V}}{dt}$) In a continuous setting, this would be the time derivative of $\mathcal{V}$, but here it is simply a running difference of the variables $\mathcal{V}$ over time.

($S : \mathcal{U} \to \mathbb{R}$) A measure of the Return on Investment (ROI) of a given institution

We will also be interested in a "sliding window" of times trailing a given time; if we're interested in a subspace of time $T \subset \mathcal{T}$, we can construct this window as $\mathcal{W}_T : T \to \mathcal{D}$. Note that in the case of a discrete T, comprised of n time steps, $\mathcal{W}_T = \mathcal{D}^n$. All of our data is discrete, of course, but this construction also works for continuous subspaces of $\mathcal{T}$, which means we can think of different choices of T as appropriate approximations, instead of sets with fundamentally different structures. It is also useful to include a shorthand several particularly important parameters

($\mathcal{M} \subset \mathcal{D}$) The space of donations. This is a special subset of $\mathcal{D}$, as we are constrained by a total sum of money, and trying specifically to optimize with respect to this quantity.

($M_u \in \mathcal{M}$) The total amount of money given to university u over the entire time T

With this framework, we can now formulate the problem more precisely. To do any kind of induction at all, it is necessary to make some commonplace but sometimes very wrong independence assumptions. Here's our first and most central one: we will assume that the effectiveness with which an institution can use money does not change over time[1]. For an institution $u \in \mathcal{U}$, we can now talk about the effect of donor money over time on the variables in $\mathcal{V}$ as a mapping

$$F_u : \mathcal{W}_T \to \frac{d\mathcal{V}}{dt} \tag{1}$$

That is – a function which takes the financial input we control, and changes student metric variables in some way. This is to be distinguished from some other function $\phi_u : \mathcal{W}_T \to \mathcal{V}$ that estimates the metric variables directly. There are several reasons for this. First of all, it makes a lot more semantic sense; we can't directly impact the absolute value of the these variables with money; at best, we nudge them in one direction or another. Secondly, training on the raw values incentivizes memorizing institutions' particular statistics instead of learning a general pattern. Speaking of which, to get this kind of general pattern instead of one for each school, we are actually learning an alternate formulation of (1) –

$$F : \mathcal{W}_T \times \mathcal{U} \to d\mathcal{V}$$

Supposing that we have a suitably robust approximation for F, we still have a ways to go in terms of allocating resources. The biggest, most glaring issue is that we still haven't decided what makes a given change $d\mathcal{V}$ a *positive* one. The mechanism for determining this will be discussed at length later in the paper, but here we will just give it a name. Let

$$A : \mathcal{V} \to \mathbb{R} \tag{2}$$

be an aggregation function, which produces an objective "goodness" of a given array of metric variables. Note that here we use $\mathcal{V}$, not $d\mathcal{V}$, even after making such a point of doing precisely the opposite with respect to F. We do this because, *a priori*, it is possible that schools have intrinsic

[1]This is a reasonable assumption to make; while technically invalid, it seems very natural to judge an institution by its past performance – indeed, this is the best we can hope for from a dataset

properties that impact the effectiveness of donations, but which do not actually themselves change with respect donation amounts. This is pretty inconvenient, because we really wanted to compose A with F to get a complete function with which to score elements of $\mathcal{D}$. For a moment, we'll suppose that we had the hypothetical function ϕ described above. In that case, we have

$$\mathcal{W}_T \xrightarrow{\phi} \mathcal{V} \xrightarrow{A} \mathbb{R}$$

...but this still isn't exactly the metric we're after. We have some absolute value of a school with a given donation distribution $w_T \in W_T$, but we'd really like to look at how quickly the composite function $(A \circ \phi)$ changes with respect to additional donations. Replacing ϕ with its description in terms of its differential, we obtain a scoring algorithm that looks like this:

$$S(u) = \frac{d}{dM_u} A\left(V_u + \int_T F(w_T, u)\ dt\right) \tag{3}$$

where $V_u \in \mathcal{V}$ is the most current metric vector for university u. Now, we can do some multi-variable calculus magic to get something easier to calculate.

(1) Apply Machine Learning techniques to estimate the effect of donations on the

4. Data Collection and Variable Selection

To carry out our analysis, we used two datasets. The first was the Delta Cost Project database[2], a dataset derived from the IPEDS datasets containing time series data from the academic years 1986-87 and 2011-12. The second was the College Scorecard dataset,[3] which contains data by year from 1996 to 2013 on a variety of measures of student performance, earnings and debt. As discussed, our goal was to determine the influence of certain financial variables on measures of student success. To this extent, we trained several machine learning classifiers with the financial data from the Delta Cost Project as our features, in an attempt to predict the differentials of student success, calculated from the College Scorecard dataset. We used two classifiers, and the approach to variable selection and preprocessing differed between the two.

4.1. Support Vector Machines.

Suppose we have a set $\{f_i^{(t)}\}_{i=1}^n$ of financial variables drawn from the Delta Cost Project database, and a set $\{s_i^{(t)}\}_{i=1}^m$ of measures of student success. For example, one of these may be the amount of private contributions to a university during the year t. Our approach here was to train an SVM regressor for each $s_i^{(t)}$, using the $f_j^{(t)}$ as features.

4.1.1. *Variable Selection.*

In order to use this work to help decide how to distribute money of the Goodgrant Foundation, we chose as features those variables over which we had direct control. For instance, we chose a variable from the Delta Cost Project database that represented the amount of revenue from private gifts, grants and contracts. We also picked as features the variables that represented the amount of revenue spent on research, student instruction, and so on. A complete list of variables chosen, for both features and targets, can be found in the appendix.

In addition, to account for the fact that different schools may handle revenue more or less effectively, we used a one-hot encoding of the school ID and added those to our $f_i^{(t)}$. More precisely, if there are n distinct school IDs, then we expand the school ID feature into n single bit features. From now on, we refer to this as the *one-hot encoding* of the feature vectors.

[2]Provided at https://nces.ed.gov/ipeds/deltacostproject/.

[3]Provided at https://collegescorecard.ed.gov/data/.

4.1.2. *Data Preprocessing.* In order to account for the temporality of the data, we used a window of five years when training and evaluating the SVM regressor. If we let $X_u^{(t)}, Y_u^{(t)}$ denote the financial and student success data, respectively, this means that for each school u, and for each year $t \in \{1996, ..., 2014\}$, we formed the single example $(h(X_u^{(t)}), Y_u^{(t)} - Y_u^{(t-1)})$, where

$$h(X_u^{(t)}) = \text{concat}(X_u^{t-1}, ..., X_u^{t-5}),$$

and the concat operator represents the concatenation of these row vectors.

Due to computational constraints, we were forced to retain only a single copy of the one-hot encoding in each concatenation (as opposed to 5, the size of the window). While this may also seem like a quite natural decision, this can affect the complexity of the model being learned. While normally the SVM regressor would have had to learn from the data that the same weights are relevant to each of the five one-hot encodings, we have essentially given it free without paying for it by more training examples.

Finally, we have also dealt with the issue of missing or not-a-number feature data by using the mean strategy for data imputation.

4.2. **Recurrent Neural Networks.** Since our data is quite naturally sequential, and recurrent neural networks have had immense success with sequential data, we tried using them for our analysis. In addition, in a recent paper **??**, an approach is described to do machine translation using Gated Recurrent Units in the form of an encoder and a decoder. The encoder essentially provides a summary of the data, which the decoder attempts to recover.

As described before, our approach consists of several steps, one goal of which is to produce a ranking function. We thought to use these encoder-decoder RNNs since they could essentially learn a single number summary of how the success data was influenced by the financial data. The hope is this is that these single number summaries, perhaps transformed by a softmax, would encode some sort of ranking.

5. Hello

Other things